Mamba Perceptron

Trevor Davis Ducharme

October 9, 2024

1 Introduction

This document provides a mathematical overview of the Mamba perceptron model.

2 Perceptron Model

A perceptron is a classifier that generates its weights based on the previous generated weights. The output of the perceptron is given by:

$$Z(X, W, B) = \left[\left(\sum_{i=1}^{n} W_{j,i} \cdot X_i \right) + B_j, \ldots, \left(\sum_{i=1}^{n} W_{m,i} \cdot X_i \right) + B_m \right]$$

$$gx = [x_1, x_2, \ldots, x_n, pw_{1,1}, pw_{1,2}, \ldots, pw_m, n]$$
$$w = S(Z(gx, gw, gb))$$

$$z = \left(\sum_{i=1}^{n} w_i \cdot x_i \right) + b$$
$$y = A(z)$$

where:

- **A** is the activation function
- **S** is the softmax function
- **gw** is the generator weight vector
- **pw** is the previous weight vector
- **x** is the input vector
- **gx** is the input vector with the previous generated weights appended
- **b** is the bias term
- **gb** is the bias terms for the generator layer

3 Training the Perceptron

The perceptron is trained using the following equations:

3.1 Necessary Derivatives

$$\frac{\partial y}{\partial z} = A'(z)$$

$$\frac{\partial z}{\partial w_i} = \mathbf{x_i}$$

$$\frac{\partial w_i}{\partial g w_i} = TODO$$

$$\frac{\partial z}{\partial b} = 1$$

3.2 Updating the Weights

The weights are updated as follows:

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

where:

$$\Delta \mathbf{w_i} = -\eta \cdot L'(y, \hat{y}) \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial \mathbf{w_i}}$$

Here, η is the learning rate, L is the loss function, and $\hat{y}$ is the target output.

3.3 Updating the Bias

The bias term is updated as follows:

$$\mathbf{b} \leftarrow \mathbf{b} + \Delta \mathbf{b}$$

where:

$$\Delta \mathbf{b} = -\eta \cdot L'(y, \hat{y}) \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial b}$$